Working with Roman CGI Simulated Datasets

Marie Ygouf – Technologist
(she/her/hers)
Jet Propulsion Laboratory, California Institute of Technology

Neil Zimmerman, Vanessa Bailey

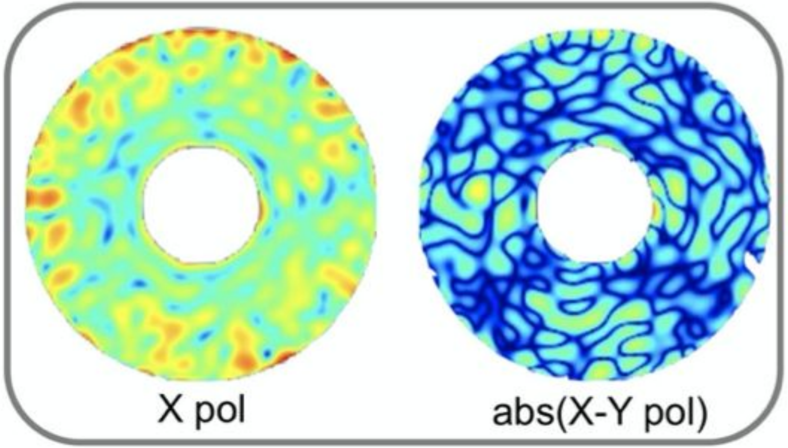
Jet Propulsion Laboratory
California Institute of Technology

Outline

- How is CGI making data processing more challenging than other instruments?
- Steps to process CGI simulated data and associated results
- Work to go and path forward to process Roman CGI images

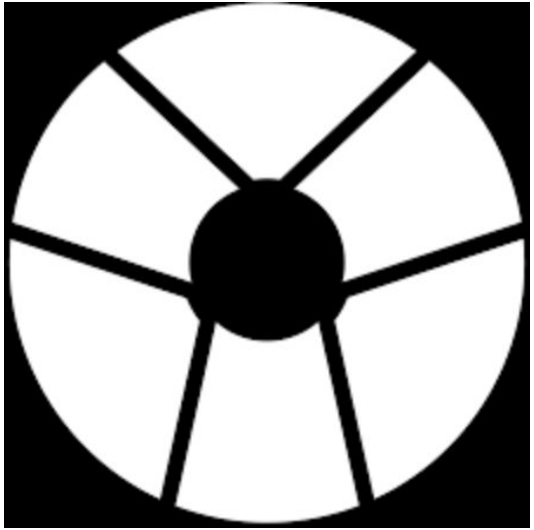
Very high-contrast imaging challenges and implications for data processing - CGI does not drive the Roman mission

- Telescope primary mirror and non-optimal mirror coatings => polarization-dependent speckles, or “polarization aberrations”



Bailey et al. 2018

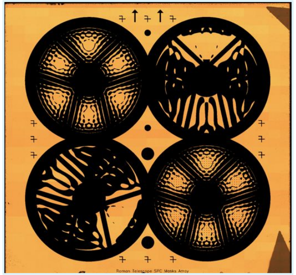
- Roman pupil not optimized for high-contrast imaging



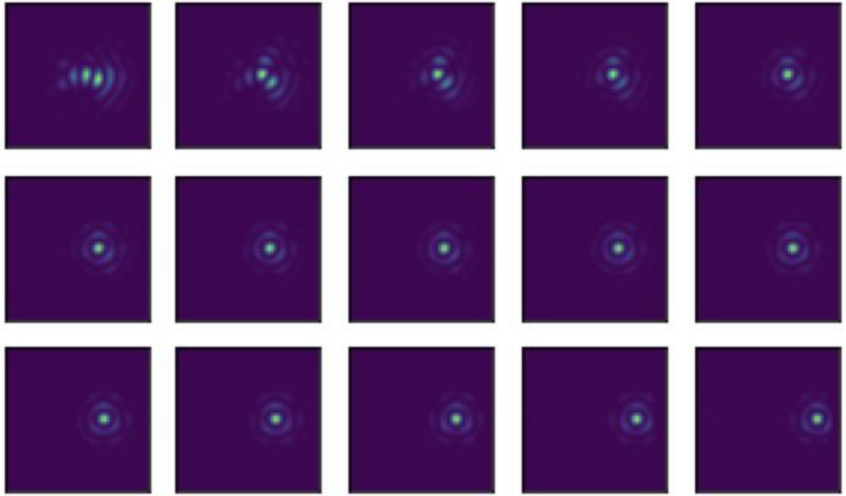
Pupil of the Roman Telescope

Very high-contrast imaging challenges and implications for data processing - Coronagraphs

- Coronagraphic mask working at very high contrast and a complex aperture, really distorted PSF.
- Implications for calibration, post-processing and analysis of Roman CGI data



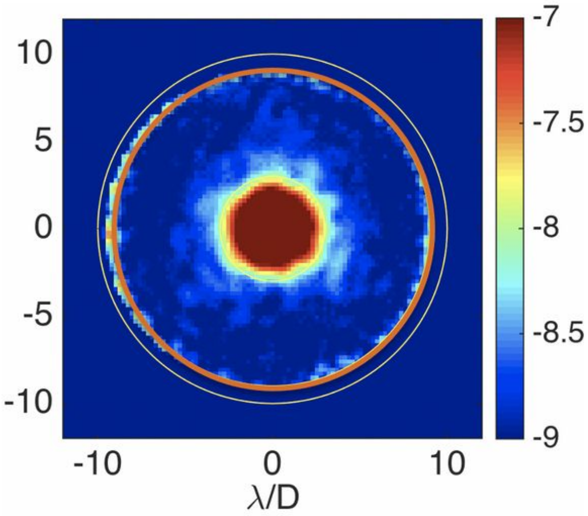
CGI coronagraphic masks



OS9 PSF for different separations from the star,
r = 1.0, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0, 5.0, 6.0, 7.0 λc/D

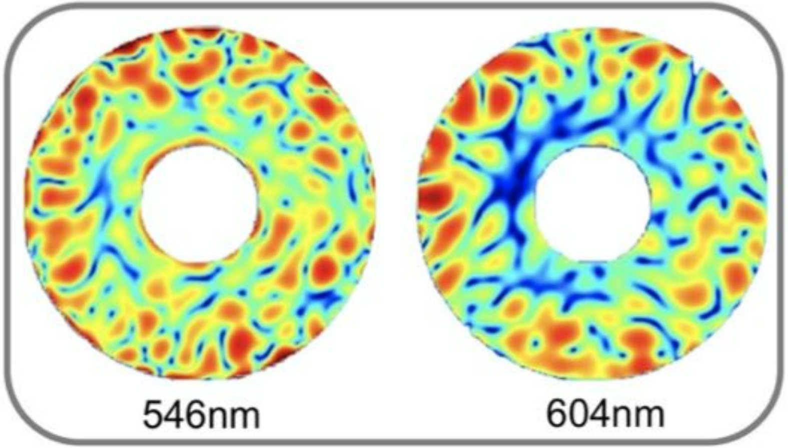
Very high-contrast imaging challenges and implications for data processing - DMs

- DMs used to correct the instrumental aberrations down to raw contrast levels of 10^{-8}



Raw experimental image with the hybrid lyot coronagraph showing the dark hole

- Implications for the behavior of speckle residuals, which is harder to predict than for standard

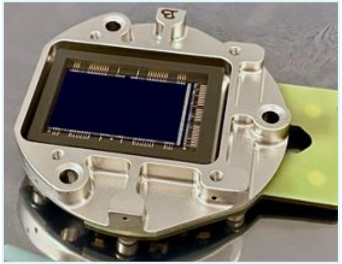


Bailey et al. 2018

Very high-contrast imaging challenges and implications for data processing - Photon counting

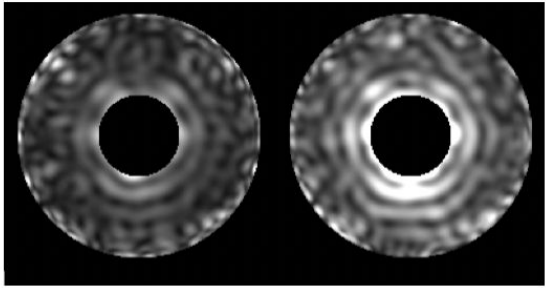
- The need for Photon Counting mode in Roman CGI
  - Cosmic rays limit single-frame integration times to a few hundred seconds and read noise is dominant (100 e- RMS per frame) => unpractical for regular CCDs
  - Roman EMCCD uses signal amplification from a gain register prior to read out to remove read noise.
  - This reduction comes at the cost of an increase (by a factor $\sqrt{2}$), called as Excess Noise Factor (ENF) of all other noises.
- Implications for post-processing
  - Photon-counting must be used to pre-process the high-gain frames in photon-counting mode

Ultra-low-noise Photon-counting EMCCDs



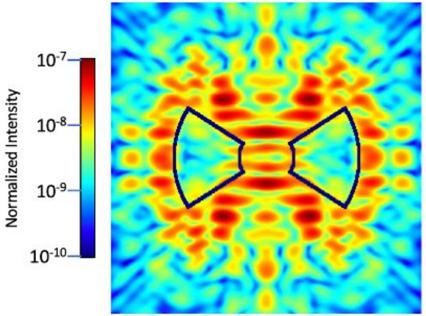
Observing Scenario 9 (OS9) Simulated Dataset

- Residual star light is the limitation of high-contrast imaging what is causing that limitation
- Generated by John Krist (JPL) - Released in May 2020
- Hybrid Lyot Coronagraph Band 1 – (Phase B design) (10% band at 575nm)
- Shaped Pupil Coronagraph Band 3 – (Phase B design) (15%, 675 – 785 nm, λc=730 nm) – Bowtie FPM



HLC - No Sensitivity MUFs

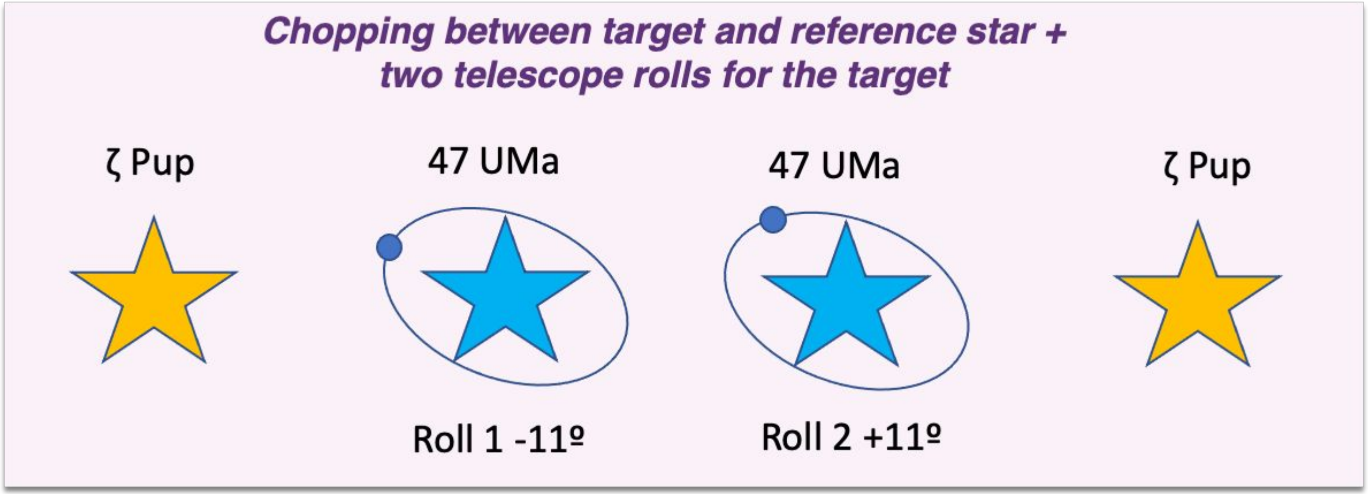
HLC - With Sensitivity MUFs



SPC FPM pattern superposed

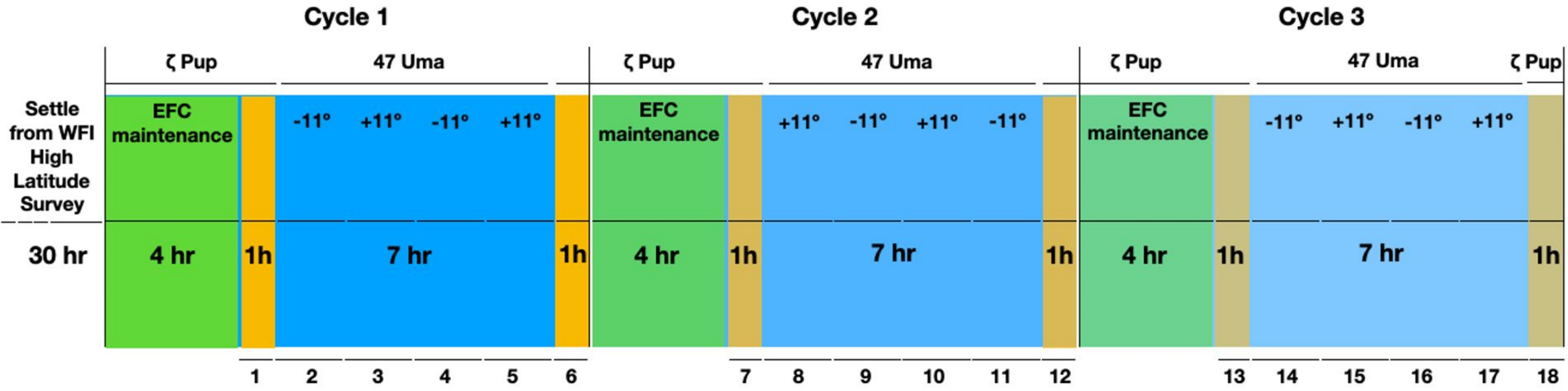
OS9 Observing Strategy

- 47 Uma (V=5.0 mag, GIV) - ζ Pup (V=2.25 mag, O4I)
- RDI & ADI (22º roll)
- Repeat observation cycle 3 times (HLC) or 14 times (SPC spectrum)



OS9 HLC Generated image time sequences

- ~20 hours on target 47 Uma, ~6 hours on reference ζ Pup
- 335 reference frames (100s each), 7080 and 7320 target star frames at roll 1 and roll 2 respectively (5s each)



OS9 HLC Datasets

File name	MUFs	Noise	Planets	Flux Units
os9_noiseless_ccd_images_no_planets.fits	-	-	-	average flux units
os9_noiseless_ccd_images_with_planets.fits	-	-	✓	average flux units
os9_ccd_images_no_planets.fits	-	✓	-	EMCCD counts
os9_ccd_images_with_planets.fits	-	✓	✓	EMCCD counts
muf_os9_noiseless_ccd_images_no_planets.fits	✓	-	-	average flux units
muf_os9_noiseless_ccd_images_with_planets.fits	✓	-	✓	average flux units
muf_os9_ccd_images_no_planets.fits	✓	✓	-	EMCCD counts
muf_os9_ccd_images_with_planets.fits	✓	✓	✓	EMCCD counts

Table from Observing Scenario 9 Post-Processing report (Ygouf et al.)

https://wfirst.ipac.caltech.edu/sims/Coronagraph_public_images.html#CGI_OS9_report

Pre-Processing

- Step 1: Data extraction
- Step 2:
  - Photon-counting procedure for ”photon-counting” mode data
  - Gain correction for analog mode data
- Step 3: Normalization

Step 1 Pre-Processing - Data extraction

- Modeling Timesteps for target 47 UMa and reference ζ Pup.
- The entire observing sequence is 25.58 hours long (excluding the EFC maintenance) and 37.58 hours long (including the EFC maintenance, not described in this table).
- Data corresponding to the EFC maintenance are not included in the OS9 distribution.

Cycle #	Target pointing #	Target	Roll angle (°)	Exp. time/frame (sec)	# of frames	Total Exp. Time (sec)
1	1	ζ Pup	-11	60	60	3600
	2	47 UMa	-11	5	1020	5100
	3	47 UMa	+11	5	1260	6300
	4	47 UMa	-11	5	1260	6300
	5	47 UMa	+11	5	1260	6300
	6	ζ Pup	+11	60	65	3900
2	7	ζ Pup	+11	60	60	3600
	8	47 UMa	+11	5	1020	5100
	9	47 UMa	-11	5	1260	6300
	10	47 UMa	+11	5	1260	6300
	11	47 UMa	-11	5	1260	6300
	12	ζ Pup	-11	60	65	3900
3	13	ζ Pup	-11	60	60	3600
	14	47 UMa	-11	5	1020	5100
	15	47 UMa	+11	5	1260	6300
	16	47 UMa	-11	5	1260	6300
	17	47 UMa	+11	5	1260	6300
	18	ζ Pup	+11	60	25	1500
				Total	14735	92100

Table from Observing Scenario 9 Post-Processing report (Ygouf et al.)

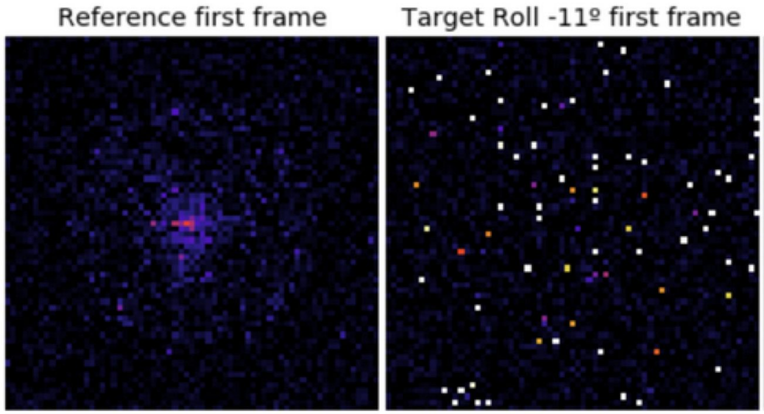
https://wfirst.ipac.caltech.edu/sims/Coronagraph_public_images.html#CGI_OS9_report

Step 2 Pre-processing - Photon counting mode

- Photon-counting mode data are high gain analog data

Analog data, 60 sec exposure time / frame

Data in photon-counting mode, 5 sec exposure time per frame



Gain = 100

Gain = 6000

Photon counting mode

- In photon counting mode any pixels with counts below the threshold are recorded as zero electrons, while any above are recorded as one electron.
- Photon Counting rejects read noise and eliminates ENF at the expense of some efficiency loss:
  - *__Threshold loss__*. This loss occurs when we record zero electrons when there actually was 1 (or more) image electrons.
  - *__Coincidence loss.__* This loss occurs when we record 1 electron, but there were in fact multiple electrons in the image pixel

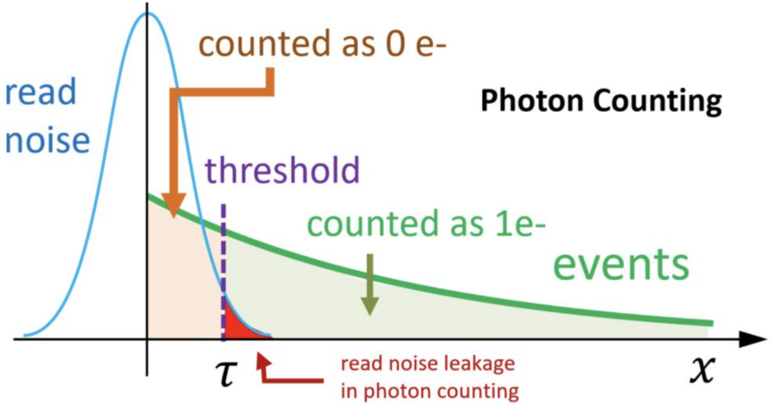
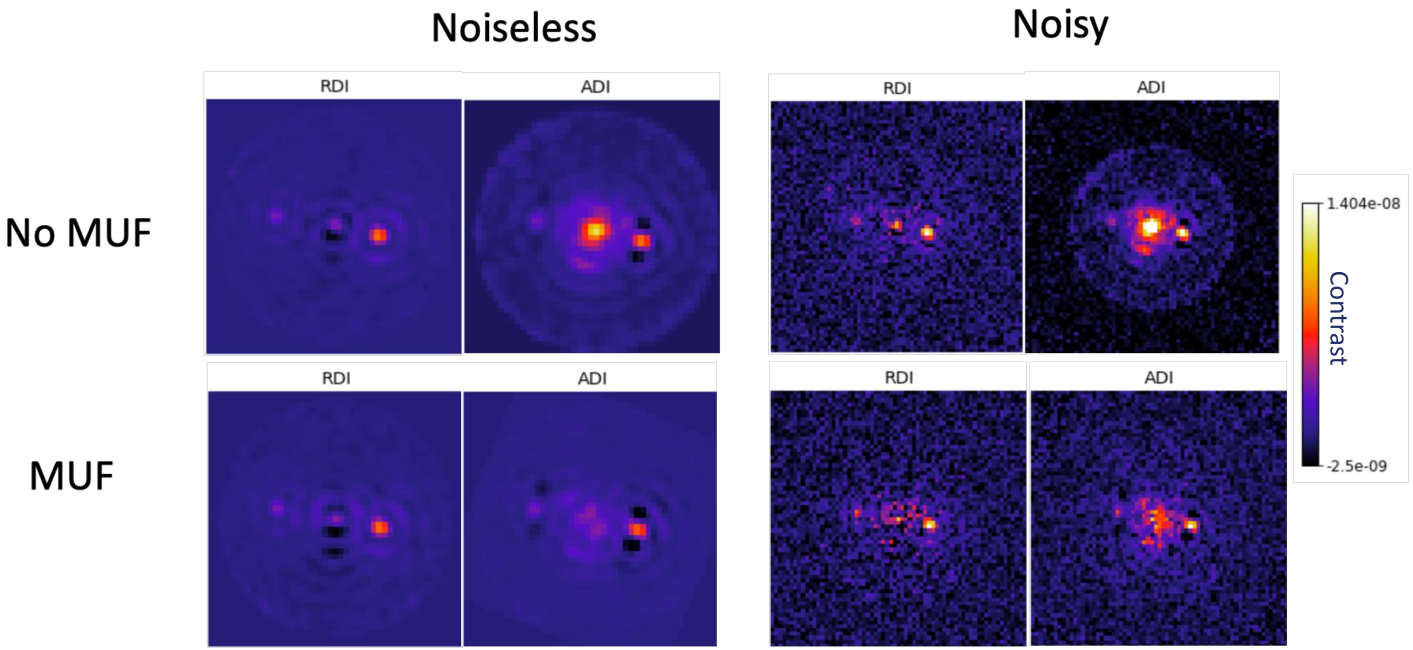
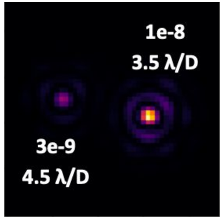


Figure from Nemati et al. 2020

- Photon counting mode procedure
  - Apply a threshold to each frame
  - Apply correction factor for thresholding and coincidence losses
  - Coadding the frames and obtain the pre-processed data

OS9 HLC data processed with classical PSF subtraction



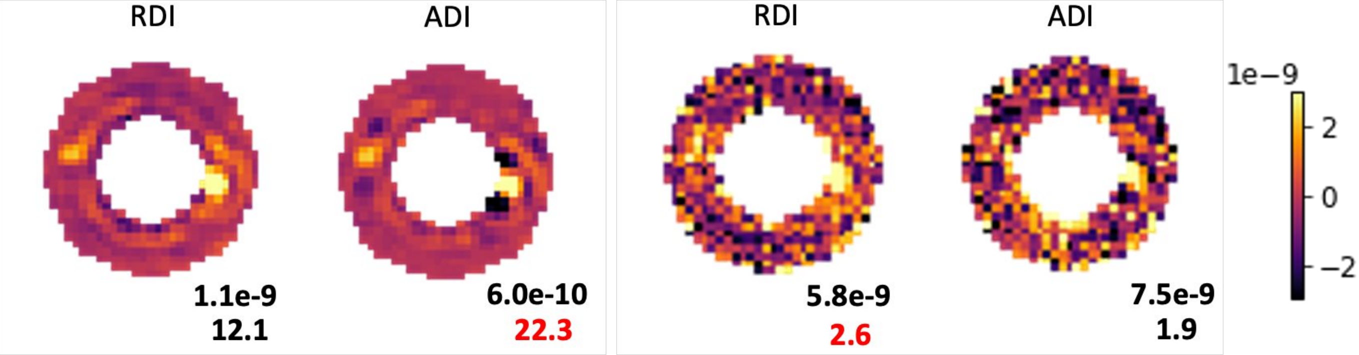


OS9 HLC data processed with classical PSF subtraction

~20 hrs on “science” target 47 UMa

Noiseless | Noisy

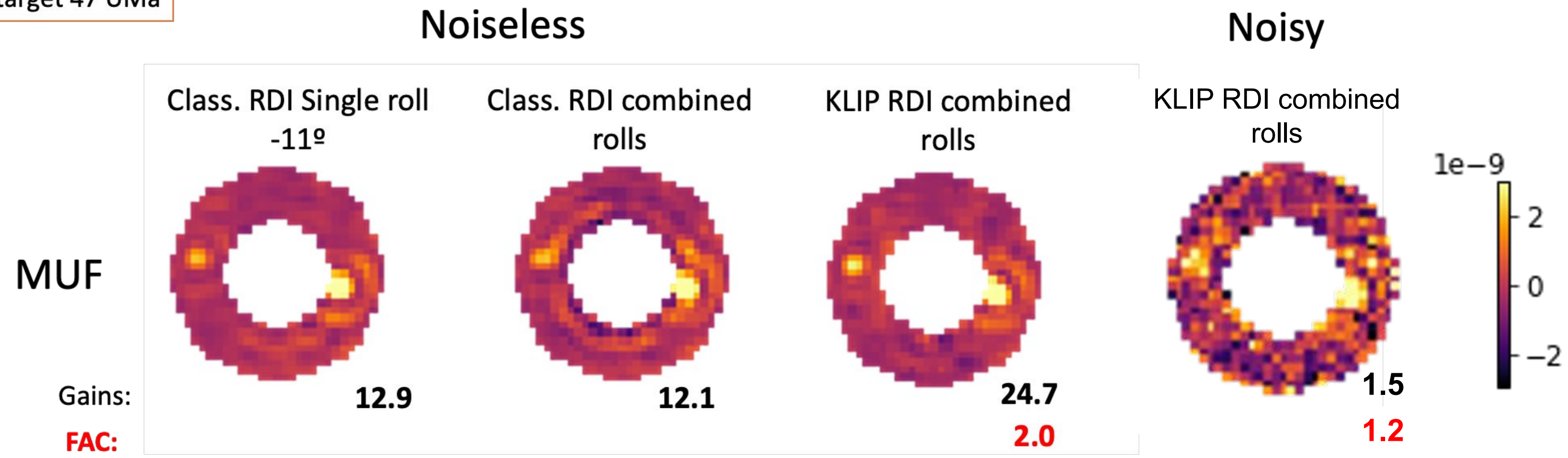
MUF



5σ contrasts:
Gains:

OS9 HLC data - Factor above classical (FAC)

~20 hrs on "science" target 47 UMa

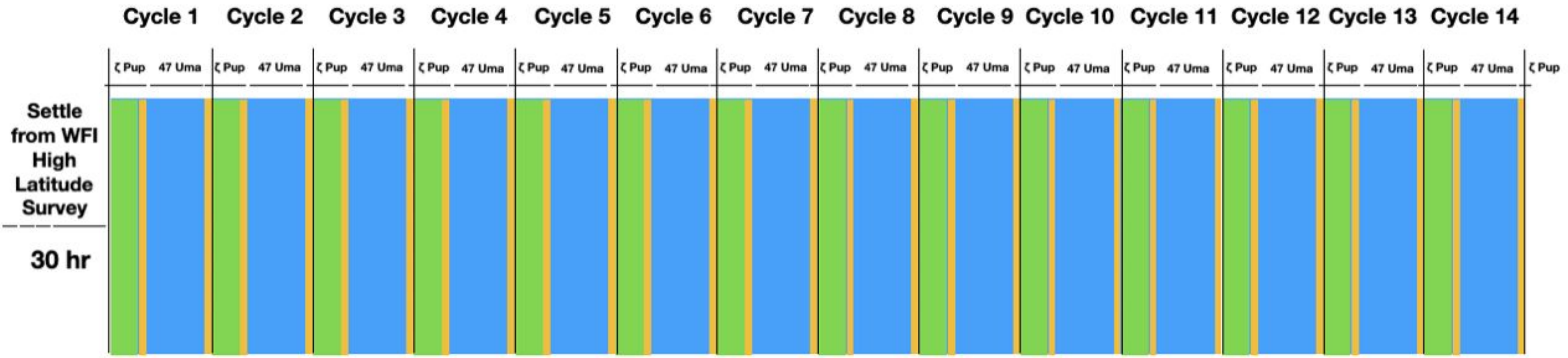


OS9 HLC data - Conclusions

- Performance of post-processing techniques on OS9 HLC data better than design requirement 10σ contrast of 5e-8
- With a total exposure time on target of only ~20 hours, noise is the limiting factor.
- Integrated gain between 3 and 5 λ/D from classical PSF subtraction ranges from ~2 to ~22 depending on the considered case
- ADI performs better in the noiseless case (speckle dominated) and RDI performs better in the noisy case (noise dominated)
- Factor above classical of 2.0 in the MUF noiseless case (speckle dominated)
- Factor above classical of 1.2 in the no MUF noisy case (noise dominated)

OS9 SPC Generated image time sequences

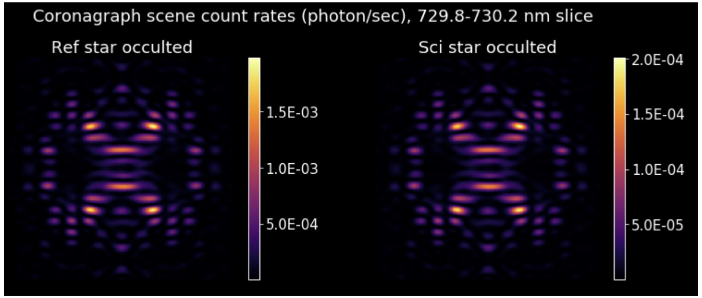
- Cycle: 4 h of maintenance EFC on ζ Pup (ζ Pup, V=2.25, O4I) / imaging on ζ Pup / slew to the target star (47 UMa, V=5.04, G1V) / 47 UMa at rolls of -11°, +11°, -11°, +11° from solar-normal roll / slew to ζ Pup / imaging on ζ Pup / Repeat observation cycle 14 times (SPC spectrum)
- No planet injected in the data



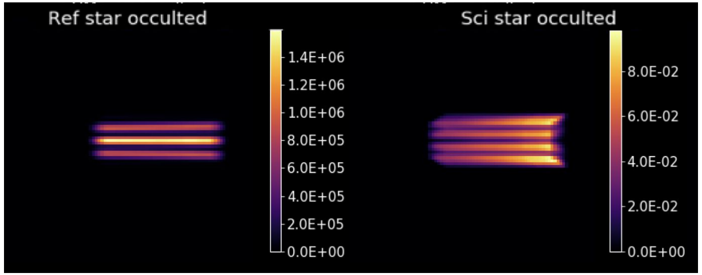
Generating OS9 SPC Band 3 Spectroscopic Data

- Goal: Compute the factor above classical for SPC Band 3
  - Used python code developed by Neil Zimmerman and Hari Subedi (GSFC)
  - Takes as input SPC OS9 simulations to create star scenes based on a specified target star apparent mag and spectral type
  - Apply the specified prism dispersion profile to the occulted science and reference stars

Before Dispersion

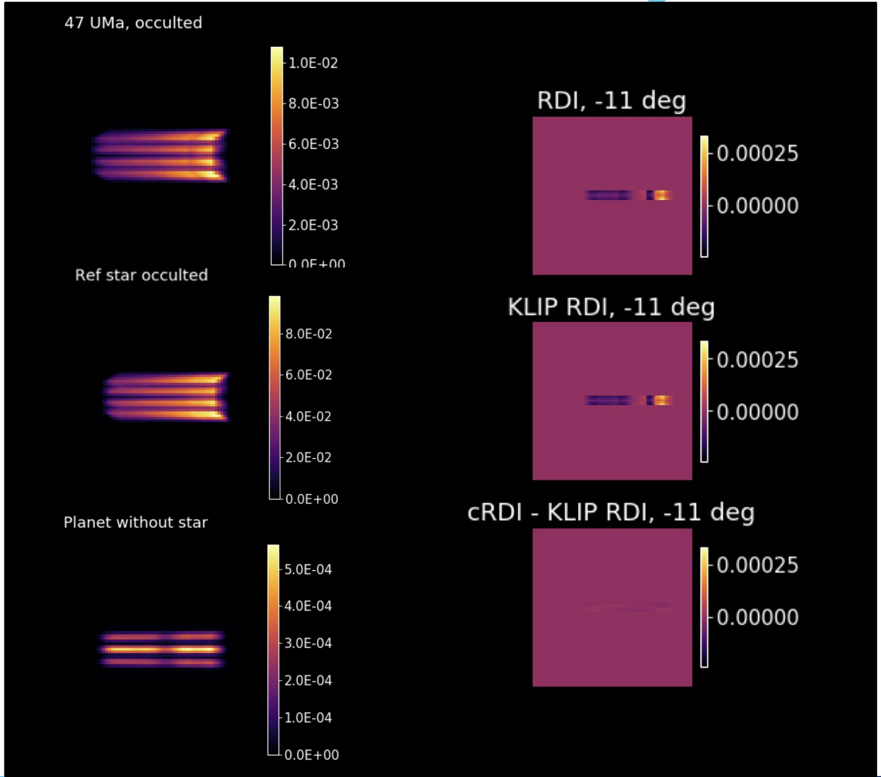


After Dispersion



Factor above classical for SPC Band 3

- Applied classical PSF subtraction (cRDI) and KLIP RDI (6 PC) on noiseless OS9 SPC spectroscopic data
- Preliminary Results:
  - KLIP throughput computed by propagating a dispersed offset PSF at the location of the planet through the KLIP algorithm
  - With this KLIP throughput taken into account, the factor above classical is 0.8 and thus cRDI performs better than KLIP RDI.
- Ideas for mitigation:
  - Select regions of the spectrum that are just above or just below the central lobes of the planet LSF
  - Take into account the presence of the planet while processing data



Conclusions

- CGI is not your typical high-contrast imaging instrument and presents some challenges for data processing
- PSF-subtractions techniques have been successfully implemented on simulated data from various OS and will be used as a baseline for the Roman CGI post-processing pipeline
- Work to go includes every aspect of data processing (pre-processing/calibration, post-processing and analysis)

Limitations and work to go

- Post-processing strategy:
  - Further optimize the post-processing parameters and regions used for the projection
  - Frame selection
- Further investigate whether PCA can improve spectroscopy results
- Post-processing of polarimetric data
- Pre-processing / calibration;
  - Develop algorithms to process calibration data
  - Develop algorithms to process data from level 2 data products to level 4 data products
- Analysis:
  - Implement photometry/astrometry using the library of PSFs (matched filter). Was done for older OS but not for OS9 and not implemented in current pipeline
  - Uncertainties estimations on photometry/astrometry (including uncertainties on the spectrum for spectroscopic data)
  - Further improve and implement useful analysis tools and performance metrics (including the use of telemetry data)

Resources

See also data simulation and processing talks by:

- John Krist - Overview of Observing Scenarios and Their Simulated Datasets
- Jessica Gersh-Range - Simulated Datasets for the "Wide" Field of View Shaped Pupil Coronagraph
- Neil Zimmerman - Spectroscopy Data Simulations
- John Debes - Disks and Exozodi: Science Case and PSF subtraction results
- Julien Girard - Exoplanet Imaging Community Data Challenge

- OS9 Simulated data:
  - https://wfirst.ipac.caltech.edu/sims/Coronagraph_public_images.html#CGI_OS9
- Observing Scenario 9 Post-Processing report:
  - https://wfirst.ipac.caltech.edu/sims/Coronagraph_public_images.html#CGI_OS9_report
- Exoplanet Data Challenge:
  - https://roman.ipac.caltech.edu/sims/Exoplanet_Data_Challenges.html
- Roman CGI parameters:
  - https://wfirst.ipac.caltech.edu/sims/Param_db.html
- Older post-processing reports:
  - OS5: Zimmerman et al., WFIRST Coronagraph Instrument post-processing algorithms for advanced PSF subtraction.pdf
  - OS5: Ygouf et al., https://www.stsci.edu/files/live/sites/www/files/home/roman/_documents/WFIRST-STScI-TR1605.pdf
  - OS1 & OS3: Ygouf et al., https://www.stsci.edu/files/live/sites/www/files/home/roman/_documents/WFIRST-STScI-TR1601A.pdf
  - Ygouf et al.,https://www.stsci.edu/files/live/sites/www/files/home/roman/_documents/WFIRST-STScI-TR1503A.pdf
- Papers:
  - "The Roman exoplanet imaging data challenge: a major community engagement effort", in SPIE Conference Series, J. Girard et al., 2020
  - "Data processing and algorithm development for the WFIRST-AFTA coronagraph", in SPIE Conference Series, M. Ygouf et al., 2016
  - "WFIRST-AFTA Coronagraphic Operations: Lessons Learned from the Hubble Space Telescope and the James Webb Space Telescope", J. H. Debes, M. Ygouf et al., , in JATIS, 2015
  - "Lessons for WFIRST CGI from ground-based high-contrast systems", in SPIE Conference Series, V. Bailey et al., 2018